

The potential of machine mediation in the maintenance of public service media

Lizzie Jackson

London South Bank University, United Kingdom

Abstract: This production study argues predictive analytics (PA) and behavioral targeting (BT) will be useful for public service media (PSM). Whereas PA is the offering of digital content based on personal data, previous browsing behaviors, and user preferences, BT is the nuanced matching of content to audience segments. The topic is of critical importance for leaders of PSM, for media regulators and for audiences. The study draws on network and new media theory to provide context. The rationale for the study is the urgent need for PSM to begin using advanced digital production processes as mediation by machine is giving commercial digital firms a market advantage. An alternative PSM approach would benefit the quality of the public sphere. The use of PA and BT is considered in four cases: the websites of *The Guardian* and the BBC in the UK, and *The Washington Post* and CNN in the USA.

Keywords: public service media, audience analysis, predictive analytics, behavioural targeting, big data, privacy.

Introduction

This production study on the distribution of digital media via computer networks explores predictive analytics (PA) and behavioral targeting (BT) also referred to as behavioural advertising; the automated ‘matching’ of content to audiences. Network Theory informs us that mediation by machine is an inherent aspect of the diffusion of content via computer networks and indicative of the recombinatory nature of digital media. PA is the collection and analysis of large reservoirs of personal and usage data (big data) by complex algorithms, often on an industrial scale. When sequentially coupled with BT – the matching of content to audience segments against previously indicated interests – ‘natural’ online browsing patterns are disrupted. It is argued that such mediation by machine is more than mere marketing, it is a way of delivering content to specific audiences. The large-scale adoption of PA and BT by digitally adept commercial firms such as Facebook, YouTube, Amazon Prime and Netflix has given them a market advantage and commodifies audiences. Machine mediation can result in a ‘filter bubble’ or disruptive narrowing of information sources that becomes critical in the case of news services (Silverstone, 2007; Pariser, 2012). The topic is of the utmost importance for public service media (PSM) being concerned with the quality of the public sphere, and also for media regulators and audiences. The rationale for the study is to assist PSM to adjust focus from the currently dominant delivery paradigms of television and radio.

In order to find out how broadcasters and newspapers are beginning to use PA and BT, content analysis of four case studies are provided: the websites of *The Guardian* and the BBC in the UK, and *The Washington Post* and CNN in the USA. The sample group offers a level of international context and a range of orientation between public service and commercial activities. The Home and main section pages

of the news sections of each websites was examined along with the website's Terms and Conditions and Privacy Policy. The data was collected April–June 2015 and updated in January 2017 with a review of the privacy policies.

The growth of automated analysis and delivery

In the 'pure broadcast' era, audience analysis largely related to the reach and popularity of television and radio programming. In the 1990s, broadcasters launched websites to accompany news channels and these began to offer the ability to examine audience consumption patterns in a more granular way through (for example) page 'views,' user retention (dwell time), website entry and exit routes and 'click-throughs' (browsing to other content). The audience analysis systems of broadcasting and Web-based media remained separate. The distribution of both broadcast *and* Web programming via internet protocols to SMART televisions, mobile devices and computers can extend this wider audience analysis to include broadcast content.

Automated mediation tools make it possible for YouTube, Netflix and Amazon Prime to understand how, when and where to deliver content to suit individual subscribers. These 'internet-born' companies have both the technical knowhow and departments of data scientists able to develop sophisticated algorithms that match people to content they might like. In the case of Amazon, the user data may already be available from previous transactions. In the 'Dot-com Boom' of the 1990's, the online retailer was able to amass tremendous amounts of data on consumers' buying preferences (Provost and Fawcett, 2013: 319). They also collected customer reviews and saw this as something that could give the company a critical commercial advantage. Skeggs and Yuill's 2013–2016 study 'Values and Value' looks at "what happens when economic value is accumulated from spheres previously considered non-economic, and the distortions and alterations in both economic and non-economic values this process creates" (Skeggs and Yuill, 2017). The researchers find "Facebook makes considerable economic gain from commodifying friendship through the algorithmic conversion of 'likes' into advertising sales" (ibid); a new form of capital they term 'emotional capitalism.'

Global giants have the finance to speculatively expand and is now turning its attention to becoming a supplier of high quality media informed by audience feedback: "Amazon.com launched fourteen pilot programmes for its streaming video service and used customer feedback on it to help it select the five it ultimately produced" (Davenport, 2014: 49). Amazon, Netflix, YouTube and others are turning themselves into media producers by commissioning rich linear media for live-streamed and on-demand distribution and this is creating an increasingly challenging climate for public service media. The understanding and use of dynamic real-time audience analysis and distribution systems such as PA and BT are essential to compete with such commercial media (Davenport, 2014, Brynjolfsson and McAfee, 2014).

It is important at this point to ensure a shared understanding of concepts and terms. The internet is framed by Manovich (2013) as a 'meta-medium' comprised of computer code and concerned with the diffusion of content within networks of computer networks. In computer science, diffusion refers to the movement of objects (such as a video or image file) through computer networks. Jenkins and others refer to the spreading of media by forwarding, recommending and so on as being an intrinsic element of participatory culture (Jenkins et al, 2013). Audiences are able to browse through Web-based recombinatory content via SMART televisions, mobile devices

and computers in consumption patterns that resemble the flocking or swarming of birds and insects depending on what is of interest at any given moment (Gloor, 2006; Miller, 2010). Web-based media is a highly fluid media landscape referred to as a 'mediascape' by Głowacki and Jackson (2014) to point out the intrinsic difference from the linear distribution patterns of broadcast media.

Network theory illustrates computer networks are relational, they depend on hyper-connectivity and on being able to pass files, messages, and commands around routes that have been identified as being useful. This spreading of material through networks is called 'diffusion,' a term widely used in computer science. Web-based delivery offers multi-directional, multi-platform, distribution and consumption. Web-delivered media is served from computer 'servers' to receiving computers and viewed via 'web browsers' using internet protocols. Browsers are user-interfaces designed for the easy viewing of content. Instagram, Vimeo, Spotify, Netflix and so on run on what industry producers term 'platforms.' Platforms are driven by sophisticated databases capable of storing and retrieving large amount of 'big data' at very high speeds. Big data refers to a variety of data, as Davenport explains "Some have defined big data in terms of three Vs (volume, variety, and velocity)" (2014: 7). The data we are concerned with here is audience browsing behaviour and personal information. Databases and their associated algorithms also offer the comparison, storage and retrieval of material to suit pre-defined preferences; a matching and responding exercise. To make these calls and responses faster, databases use algorithms. An algorithm is "a set of steps that are followed in order to solve a mathematical problem or to complete a computer process" (Merriam-Webster.com, 2016). Databases can communicate between platforms (SMART TVs, tablet computers, laptops, and mobile phones) as long as they are all using computer code. This gives 'Inter-operability'; the ability to view the same video file, photograph, or word document on different receiving devices.

I argue PSM needs to become involved in the development of machine mediation works for several reasons. First, in order to provide a counter to what Pariser calls 'The Filter Bubble' (Pariser, 2012). Both Cushion (2012) and Iyengar and Kinder (2010) argue that audiences who are served only closely related content or 'same opinion' commentary may not receive a sufficiently broad selection of news resulting in an ever-reducing understanding of our world. Secondly, techniques such as PA and BT offer a means to ensure all audiences are made aware of the availability of public service content in a more sophisticated way than merely offering content 'on demand.' Audiences can indicate preferred reception devices and other access parameters to suit their competencies, age, location and so on. Third the lack of any balancing presence of PSM in debates relating to the future use of PA and BT might have implications for the mediascape going forward. Lastly, PA could be merely framed as 'Audience Analysis 2.0' however it is much more than that as it is also enmeshed in the fabric of the delivery of Web-based media *across a variety of different* delivery platforms. BT could be thought of as mere marketing, but again, it can be seen as a highly sophisticated way to nuance content delivery to specific audiences. Machine mediation is therefore of high importance to PSM as it creates a fresh model of content distribution.

Any lack of understanding of how Web-based media is delivered, and how machine mediation works, results in commercial media having an advantage. There is currently a critical imbalance of power between PSM and commercial media that is resulting in the growing commodification of audiences (Castells, 2009; Graham and De Sabbata, 2013; Freedman, 2014). In addition there are rising concerns worldwide

relating to the misuse of personal data noted by Ess (2009), and Silverstone (2007). New European legislation in connection with the use of personal data will take effect in 2018 (European Commission, 2016).

An overview of how PA and BT works begins the chapter. This is followed by a short review of potential uses and abuses. An explanation of the international self-regulatory frameworks is then followed by a look at the four case studies, each of which is a provider or aggregator of news. Each is making the transition from being a 'traditional' media provider (broadcasting or newspapers) to a media landscape increasingly dominated by computer networks.

Defining machine mediation

Theories connected with 'New Media' show us how Web-delivered content is recombinatory (Bolter and Grusin, 2000; Lievrouw and Livingstone, 2002). Computer network theory assists us to see an emerging media landscape of multi-directional flows of material (Tanenbaum and Weatherall, 2014) that can be viewed on many different receiving devices. Networked media is a highly adaptive medium with a varying level of fluidity (Głowacki and Jackson, 2016) that 'solidifies' at times to enable reification, the making of meaning out of a thing (Vitale, 2014: 19). For example, moments of stasis can be found in the latest edition of a website home page. Computer science also gives us the notion of 'diffusion,' the spreading of material through networks of networks, a term also used by biologists to describe states of liquidity. Increasingly, the mediation of such flows of Web-based content is being automated. Text, images, video, other such media and social commentary is being diffused through networks using algorithms able to draw from large media stores in databases. It is impossible to do such processing by human hand; PA and BT are two examples of mediation by machine.

Siegel's definition of PA is "Technology that learns from experience (data) to predict the future behavior of individuals in order to drive better decisions" (Siegel, 2013: 11). For Web-based media this is the analysis of both previous browsing behaviours and personal data. PA can be augmented by BT, which is the segmentation of audiences into targetable groups. BT can be seen as a process that takes place after a user's browsing patterns and personal information have been analysed. According to the Internet Advertising Bureau (a standards body promoting self-regulation for the advertising industry) BT is comprised of *contextual advertising* and *demographic advertising*:

"contextual advertising is where advertisements are served within a chosen 'context' by the selection of a website focused on a particular topic. An example is when a user is shown an advertisement for tennis rackets solely because he/she is visiting a tennis-related website. Behavioural advertising is also different to demographic advertising where advertisements are served based upon specific information provided by the user (e.g. gender, age, location)." (The Internet Advertising Bureau, UK, 2009: 13)

An *advertising server* is a company that specializes in creating the technology that delivers advertising (banners for example) to the relevant page on a website. An *advertising network* is a firm that connects web-based businesses, content creators and advertisers. An *advertising exchange* acts as an open broker facilitating the buying and selling of advertising. A *data aggregator* is a firm aggregating information from many sources that might interest audience segments that have

indicated a propensity for a marketable commodity or service.

Inter-operability between linked databases and content management systems offers the automated triangulation of member accounts, content files (video/audio/text/animation/photo and so on), plus – increasingly – geo-locate data. All this can be used to direct selections of media to different media platforms. At best this results in more meaningful selections of information, at its worst we may no longer be sharing the same public sphere but be experiencing a constructed reality. At the very worst, we may not even know this is happening because “In cyberspace we are constantly surrounded by relational databases, often without even realizing it” (MacCormick, 2012: 147). This brings us on to the need to address the potential benefits and drawbacks of machine mediation.

Uses and abuses

The Web-based media landscape is a meta-medium; composed of meta-tagged files that can be recombined to create both existing and new forms of media (Manovich, 2013). Organizations that understand and have capacity to produce such media are able to exploit the interconnected nature of networks to produce economic and societal value plus other newer forms of capital such as that relating to the previously noted ‘emotional capitalism’ identified by Skeggs, Yuill and Zschomler (2016). Televisions, computer tablets, and mobile phones are now capable of receiving web-based content. Marketing messages and purchasing opportunities can be inserted into browsing sessions, based on predicted interests. Young people get most of their news via social media (BBC News, 15 June, 2016) therefore we argue a very strong PSM presence online has become critical. For PSM this is irrelevant if you currently have little or no Web-based services.

The BBC launched online services in 1997 and this has more recently expanded to a broadcast channel (BBC3) in 2015 in order to cater for the media preferences of 16-35 year-old audiences. They are beginning to explore the automation of mediation, as a practical necessity. Their User Generated Content (UGC) Hub launched in 2005 and the journalists staffing the UGC intake desk use – mostly manual – matching to verify such content before it is used to substantiate or augment more conventional news stories (Meier, 2015 :135). The volume of UGC received by the intake desk is high. There’s a section on the BBC’s news app for mobile phones (‘My News’) that responds to previously chosen interests an indication of the use of dynamic audience analytics. The BBC’s News Lab is also experimenting with text-to-speech and artificial intelligence with a trial News Bot for BBC World and BBC Mundo. The bot will enable the sharing of news stories with others, plus suggest a new batch of links at the same time. One of the developers of the service, Labber Lei He says the BBC wants to keep the service simple and transparent in order to counter “concern over black box algorithms and biased distribution platforms, so we wanted to give users clear, simple commands to control the news they receive” (BBC News Labs, 2017). The BBC is also experimenting with new services that use cookies and other Web-based audience engagement processes to enable interactive and social platforms (<http://www.bbc.co.uk/taster/>).

The Japanese public service broadcaster, NHK, has recently been developing an earthquake and tsunami alert and response service for the public that aggregates sources of big data connected with such events. These trials have been successful and the service is likely to expand. The Canadian Broadcasting Company (CBC) has been adjusting its newsroom to enable the automated ingest, retrieval and delivery of news

to multiple devices. The Head of CBC News Strategy, Andrew Cochran, envisions ‘a river of content’ that can be automatically sifted and received by any device to suit the requirements of CBC and the public (Cochran, 2016).

Machine mediation has other uses that audience analysis; it can produce new forms of expression such as data journalism. Hurrell and Leimdorfer (2015) define data journalism as broadly covering projects that use data to do one or more of the following:

1. Enable a reader to discover information that is personally relevant,
2. Reveal a story that is remarkable and previously unknown,
3. Help the reader to better understand a complex issue.

The researchers argue that data journalism has some longevity; the BBC has been offering simple tools for audiences to use (for example to interrogate finance and weather data) for over ten years. Simon Rogers who pioneered the development of data journalism when he was at *The Guardian* believes the concept is even older citing Florence Nightingale’s use of graphical data to articulate mortality rates in the army from 1858. Data journalism “incorporates such a wide range now of styles – from visualization to long-form articles. The key thing they have in common is that they are based on numbers and statistics – and that they should aim to get a ‘story’ from that data” (Rogers, 2015:5). Rogers believes good data journalism is connected to drawing on diverse sources, a perspective supported by a range of scholars (Silverstone, 2007; Pariser, 2012; Freedman, 2014).

We now explore the characteristics connected with machine mediation that may have potential negative outcomes. One aspect referred to earlier is the narrowing of news caused by the ‘filter bubble’ offering restricted news sources based on your previous choices or opinions you concur with versus the challenging and the surprising. The matching that computers are able to do is becoming less crude; one of the essential characteristics of computers is they are able to ‘learn,’ sometimes referred to as ‘machine learning.’ Machine learning has been most prominently demonstrated by IBM’s super-computer Watson that, in 2011, challenged a US game show Jeopardy! to a game of chess (BBC, 2011). Watson performed better as time went on, as the machine began to guess the probability of the game more accurately. According to McKinsey “machine learning is nothing like learning in the human sense (yet). But what it already does extraordinarily well—and will get better at—is relentlessly chewing through any amount of data and every combination of variables” (McKinsey, 2015: 2).

PA and BT are likely to assist producers to deliver content in a more nuanced way, as has been argued, however there is still a need for human intervention, for example in making editorial judgements connected with prior knowledge, cultural nuances, and so on. It is also still necessary to programme the ‘learning machines’ in order to have an acceptable level of consistency, repeatability, and objectiveness (Provost and Fawcett, 2013: 26). Facebook has recently been concerned with how to manage the injection of ‘Fake News’ items into its services (Solon and Wong: 16 December, 2016) and this has resulted in the forwarding of potential fake news to external – manual – fact-checking organizations such as Factcheck.org.

One of the key issues relating to the use of machine mediation is that there are fewer non-commercial platforms serving quality news using dynamic database-driven networked systems. Young people use social media platforms to learn about their world and the lack of balanced arrays of news is worrying. Silverstone (2007) and Cushion (2012) argue that being in possession of the facts as they are known from a

sufficiently diverse range of sources is key to the preservation of deliberative democracy. Alex Pentland, a prominent data scientist also expressed his concern at a lecture hosted by the UK news aggregator *The Guardian* in January 2015:

“(…) the formation of social media echo chambers, in which a very limited number of viewpoints are given precedence and perpetually reiterated on sites such as Twitter. Pentland says that so many voices are given a platform that users become unable to assess the value of each and end up following a tiny number.” (Williams, 2015)

Objectivity and ensuring the quality of news has long been the province of PSM, and particularly so in a news environment. Any absence of a public service presence able to act as a balancing agent will be keenly felt in the future. PSM is currently less likely to have the infrastructure to manage such complex data gathering. The kinds of people needed to run big data initiatives (i.e. data scientists) are also different from conventional audience analysts “They are more facile with data, more experimental, and more product focused” (Davenport, 2014: 15). Many PSM firms are concentrating on making the transition from being ‘pure broadcasters’ to producing content on several platforms. However, the decision not to explore data management techniques would be foolish as machine mediation offers a powerful and sophisticated distribution tool. We now turn to review the self-regulatory frameworks for PA and BT that are in place in North America and Europe.

The self-regulatory frameworks

PSM has historically had a critical voice within debates connected with the development of policy and practice relating to media and communications. I argue that PSM should continue to have a presence in such discussions. There is also an opportunity to perform an important public awareness function, but until that happens national and regional advertising standards organizations are providing information and educational material for the public alongside acting as quasi-regulators, however such material is unlikely to reach the attention of young people (high users of web-based media) and other potentially vulnerable groups.

The European Interactive Digital Advertising Alliance (EDAA) provides an educational website giving information on how behavioural advertising works and an overview of European standards to assist consumers¹. Seven key principles have emerged, of which four are noted here. First, there should be transparency in how automated tracking is used by media firms. Secondly, users should have choice and control of their own data, which should also be stored securely. Third, the monitoring and segmentation of audiences should have limitations, for example in connection with children’s browsing behaviours. Lastly the EDAA recommends that education be provided for consumers about behavioural advertising and the European self-regulatory Framework.

In the USA, the Digital Advertising Alliance (DAA), an independent non-profit organization representing American advertising and marketing associations, offers another similar self-regulatory Framework. Like their European counterpart, the DAA has produced a set of principles connected with the transparency of ‘interest-based’ advertising. They encourage media firms to inform consumers about

¹ *Your Online Choices. A Guide to Online Behavioural Advertising*. Retrieved March 8, 2016 from <http://www.youronlinechoices.com/uk/about-behavioural-advertising>

the practices and processes being used such as how to opt out. The DAA has also developed principles for mobile advertising (2013) in response to the growing use of PA relating to mobile devices. This includes the use of “cross-app data, precise location data and personal directory data.” Enforcement of these standards began in 2015 (The Digital Advertising Alliance, 2015).

Advertising Standards Canada (Adstandards.com, 2016) offers links to further information and an online tool that makes it easy for the public to opt out of tracking programmes. They also provide a ‘TM Ad Choices’ icon that is also becoming widely used in North America and Europe. The icon indicates companies who comply with recommendations on transparency and control over interest-based ads.

In order to illustrate how other news outlets are experimenting with machine mediation we now look at the four case studies, the BBC, *The Guardian*, CNN, and *The Washington Post*.

The mechanical mediation of news in the case studies

A comparative review of the use of PA and BT in news content was undertaken in the cases of *The Guardian* and the BBC in the UK, and *The Washington Post* and CNN in the USA. The Home and main section pages of each of the websites, the Terms and Conditions, and the Privacy Policies were all analyzed. The content analysis took place between April to June 2015 and the Privacy Policies for each website were examined again in January 2017. Overall, there is an increasing range of personal data being collected, this may be due to an expansion in the services these companies are offering resulting in the need for a wider range of personal data, for example, game play and game scores. There is also an increase in the embedding of third party services such as social media channels, inevitably accompanied by disclaimers and alerts advising users to check the privacy policies of these third party providers. The more commercially-oriented firms such as *The Washington Post* and CNN are making greater use of audience data analytics than the more public oriented enterprises the BBC and *The Guardian*. The BBC appears not to be updating its Privacy Policy as comprehensively as *The Guardian*, *The Washington Post* and CNN who all updated their documentation in 2016. The BBC’s web page on Cookies mentions a ‘recent’ (2012) European law connected with the use of Cookies.

Table 1 illustrates the overall range of personal information that is being collected.

Table 1: The range of personal data collected by the case studies

Item	The Guardian	BBC	Washington Post	CNN
Name	✓	✓	✓	✓
Email	✓	✓	✓	✓
Postal Address	✓		✓	✓
Telephone Number				✓
Fax Number				✓
IP Address		✓	✓	
Credit Card Number			○	✓
Geo-location via mobile			✓	✓
Date of Birth			✓	○
Gender			✓	○

Job Title			✓	
Professional Networks	✓		✓	
<i>Photo</i>				○
<i>Third Party Social Media Profile</i>			○	○

*identified in the January 2017 data update.
Source: Author.

The *Washington Post* and CNN (commercially-oriented) are collecting more personal data than *The Guardian* and the BBC (public oriented) and a wider range of data. All the firms say they comply with the self-regulatory recommendations of their nation or region. It is alarming how much personal data is being harvested. Requests to gain access to users' external social media profiles indicates an urgent need for intervention. *The Washington Post* and CNN are also tracking users using location systems such as GPS tracking via mobile phones. The tracking of audience movements and location is also being achieved through the use of cookies and beacons (Table 2).

Table 2: Tracking devices used by the case studies

Item	The Guardian	BBC	Washington Post	CNN
Cookies	✓	✓	✓	✓
Beacons				✓
Beacons			○	✓

*identified in the January 2017 data update.
Source: Author.

Cookies are small applications placed on a user's computer by a web server that collect user-preferences and enable ecommerce, such as maintaining the persistence of a shopping cart's contents. Cookies can also be used to store user data between sessions. A definition of beacons and their use is provided by Stuart (2014):

“A beacon is a low energy Bluetooth chipset known as Bluetooth Smart which has its own integrated power source and emits a 2.4GHz signal which communicates with a smartphone to pinpoint your position to within 200 feet. The big opportunity for marketers is the ability to combine the location information derived from beacons with other customer information to generate personalised offers and communications, enhancing the customer experience and maximising marketing ROI.” (Stuart, 2014)

Beacons are beginning to be used to deliver location-based advertising, for example they can be placed on a shop doorway to send an alert to a customer's mobile phone that their favourite jeans are on sale. The beacon can also provide additional information to guide the person to a specific section of a shop. The collection of the movements of users reflects the growing exchange of data between mobile devices and computing embedded in objects, referred to as the 'Internet of Things' (IoT). The classic illustrative example of the IoT is the fridge that informs you when it is empty.

Having considered *the collection* of personal data we now move on to look at *how* personal data is being used (Table 3).

Table 3: The purpose of the personal data collection by the case studies

Item	The Guardian	BBC	Washington Post	CNN
Personalization	✓	✓		
Enable participation	✓	✓	✓	✓
Contact users	✓	✓	✓	✓
Obtain user data in order to improve services	✓	✓		
Understand how services are being used	✓	✓		
Identify opportunities for selling advertising	✓		✓	✓
Market services to users	✓		✓	✓
Target adverts			✓	✓
Supply products				✓
Notify users of offers			✓	
Obtain permission to act for other agents			✓	
Subscription purposes				✓
Enable social networking				✓

Source: Author.

The declared uses of personal data appear to be far narrower than is evidenced. This is particularly true of *The Washington Post* and CNN. *The Guardian* and the BBC are collecting data in order to improve and personalize services whereas their US counterparts are prioritizing marketing, commerce, and conversion activities. One of the most worrying trends is the industry acceptance that it is the user's own responsibility to check the privacy policies for any embedded third party services. This is partly mitigated by an increase between June 2015 and January 2017 in the amount of information being provided for the public on their websites. All the organizations assume a level of media literacy and some understanding of the law. Yet, privacy policies are written for producers rather than audiences. Certainly young people, who are arguably most at risk, would find them impenetrable. The final Table illustrates how personal information is used by the firms. As in Table 1, CNN has expanded its use of PA and BT since June 2015.

Table 4: Third party use of personal data by the case studies

Item	The Guardian	BBC	Washington Post	CNN
Share personal data within the organization, group, or sub-contractors	✓	✓	✓	✓

Share data with others, with consent			✓	
May collect anonymised user data				✓
Collect data from public sources on users			✓	○
Not responsible for the practices of third parties			✓	○
Share data without prior consent.				○

*identified in the January 2017 data update.

Source: Author.

In all the cases, the firms – whether they are oriented towards public service or commerce – are trading access to more customized services in return for sequential requests for personal information. Alarmingly, CNN state that if they are acquired by another business, users’ personal details may be transferred along with other company equity.

The Washington Post and CNN are more likely to share both data with others and a wider range of data; for CNN this had expanded by January 2017 to include personal game scores. From the Terms and Conditions it is not clear whom *The Washington Post* are sharing data with, conversely Time Warner – CNN’s parent company – provides a detailed list of the companies within the conglomerate.

There is a growing use of the disclaimer in this self-regulated landscape. In 2015, *The Washington Post* declared it could not be held responsible for any embedded third party service. In January 2017, CNN also adopted this position, making it clear that any user accessing services from outside the US does so at their own risk. In 2015, CNN noted that they are a global company therefore “it is necessary to transfer your information internationally (...) The data protection and other laws of other countries, such as the United States, may not be as comprehensive as those in your country” (CNN, 2015). In 2017, CNN’s Privacy Policy now takes a more emphatic stance stating “If you are located outside of the US, you use this service at your own risk” (CNN, 2017).

The aim of this study was to find out whether PA and BT might be useful for public service media. Yes, undoubtedly this is the case, particularly if mediation by machine is framed as being a fresh distribution method. However, there are challenges that come with its use, firstly ensuring the privacy of the public is taken into consideration, second – and of no small consideration - there will be a need for investment in technology and training. To summarise the overall findings:

1. Machine mediation can deliver content to the right platform at the right time to suit users’ preferences.
2. Machine mediation disrupts ‘natural’ browsing patterns and can result in a narrowing of information returned creating knowledge silos.
3. Data journalism can offer a widening of information around topics.
4. The use of personal data is often not clear, particularly agreeing to the cross-referencing – and even collection of data - from within users’ social media accounts.
5. There is a strong need for media outlets to encourage literacy in the use of personal data using audience-specific forms of address.

6. The public is commodified and their personal data is reduced to ‘assets’ that can be acquired.
7. North American and European self-regulatory frameworks appear to be similar.

Machine mediation is growing and this is partly due to the ability of computers to learn (machine learning) and the ease with which they can sort, sift, and deliver content to identified locations. Having experienced Google and Wikipedia’s ability to return highly accurate searches, and perhaps not knowing a world before the internet, young people *expect* their media to be on the right platform at the right time, on demand. They may accept they have to trade personal information in exchange for more personalized services, or simply be unaware of the implications. Who is collecting what and for which purpose is not at all clear, which suggests a need for public service intervention at policy and industry levels. Furthermore, there is a part for PSM to play in the future development of such technologies and practices as they are becoming an intrinsic part of digital media distribution.

We have been looking at PA and BT, acknowledging that these activities have a wider function than the audience analytics of broadcasting. To demonstrate the growing enmeshing of audience data with the mechanics of media delivery see Table 5.

Table 5: Audience analysis and the delivery of digital media

Media Type	Audience Analytics	Use
<i>Broadcast Channels</i>	<ul style="list-style-type: none"> • Audience Ratings • Audience Share • Audience Reach 	<ul style="list-style-type: none"> • To judge the popularity of individual programmes and channels. • Adjust scheduling to maximize audiences. • Analyse regional location.
<i>Web-based Media</i> (Websites and also media delivered to SMART TV and mobile devices via internet Protocols).	<ul style="list-style-type: none"> • Personal Details • Browsing Patterns • Browsing Preferences • Purchasing Details • Location 	<ul style="list-style-type: none"> • Track user-engagement into/out of and between websites/Web-based media. • Define audience interests. • Group audiences into similar interest ‘segments’. • Supply appropriate media to suitable platforms. • Offer ecommerce. • Define user location. • Collect user location data.
<i>Towards Media Embedded in Objects</i> (The ‘Internet of Things’).	<ul style="list-style-type: none"> • Personal Details • Browsing Patterns • Browsing Preferences • Purchasing Details • Purchasing Preferences 	<ul style="list-style-type: none"> • To judge where to send media – which platform device, or object. • To define the location of the audience. • To judge how the

	<ul style="list-style-type: none"> • Location • Pervasive user-sensing (proximity via wearable devices). 	media was received in order to nuance future delivery. <ul style="list-style-type: none"> • Monitor user's presence. • Collect user location data.
--	----------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Source: Author.

In summary, as the delivery of media becomes linked with our understanding of the audience, their devices, and their movements, we have moved far beyond analysis of sentiment and audience numbers. The high volume of data (big data) that is now indicative of the almost live nuancing of digital media production and delivery, requires mediation by machine. These processes can no longer be done by hand. So is mediation by machine good, bad, or just inevitable?

Conclusions

The 'mediascape' is extending beyond the confines of television and radio channels, and beyond websites and web-based media delivered to mobile devices towards the 'Internet of Things.' This is the delivery of media and information to web-enabled objects such as cars, household appliances and via beacons placed strategically in doorways. Media is also beginning to be offered for downloading to accompany exhibitions or historical walks via QR Codes. This – of course – challenges the very definition of media. The volume of content being diffused through the network of networks (the internet) and particularly via the browser-based World Wide Web is of such volume that we are moving into the realm of data science. Mediation by machine is a dominant and intrinsic part of web-based media. Mediation by machine is therefore inevitable.

From a positivist perspective, PA and BT offer a highly responsive solution to supplying audiences with whatever media they want, how they want it, when they want it. The ability of web-based tracking and delivery systems offer multi-way, multi-dimensional distribution. There is also the opportunity to use the 'return path' to develop deeper relationships with the public and to understand more about them as a mass, a group, or an individual. For audiences there is the opportunity to support PSM through crowd-sponsorship and crowd-thinking, if PSM enthusiastically explores the possibilities.

For PSM there is a dependency on starting up or extending existing data management departments, on data management planning, and in the acquisition of skilled data scientists and analysts. Commercial media such as Amazon, Netflix, and YouTube are industry leaders as they have had these structures in place for many years. However, there are indications that PSM is beginning to catch up. There is a high dependency on national regulators and policy makers understanding the importance of encouraging PSM to enhance and extend their web-based activity. Conversely, there is a lack of intervention by PSM in the way PA and BT is developing, and this is worrying. PSM needs to act as a counter and strong foil to commercial outlets in the media data sphere. There also continues to be debates on the use of personal data online, and this is something that should be the natural province of PSM as a protector, educator, and guardian of the citizen in the sphere of media and communications.

We have been arguing for a higher presence of PSM in digital media networks in order to counter any narrowing of news and information resulting in knowledge silos. Networked media requires public service algorithms and machine driven PSM systems that continue to challenge the known, in exactly the same way that PSM functioned in the broadcast era. Until PSM engages with data science and data systems, commercial media (e.g. Netflix, Amazon Prime and so on) will retain an advantage. Fuelled by the commercial systems offered via online trading, commercial media are able to plough these funds into becoming purveyors of high quality media.

References

- Adstandards.com (2016). *ASC Online Behavioural Advertising Compliance Procedure*. Retrieved January 1, 2017 from <https://adstandards.com/en/OBA/index.html>.
- BBC (2011). *IBM's Watson supercomputer crowned Jeopardy king*. Retrieved January 1, 2017 from <http://www.bbc.co.uk/news/technology-12491688>.
- BBC News (2015). *Social Media 'Outstrips TV' as News Source for Young People*. Retrieved January 10, 2017 from <http://www.bbc.co.uk/news/uk-36528256>.
- BBC News Labs (2017). *Behind the BBC's Messenger News Bot*. Retrieved January 14, 2017 from <http://bbcnewslabs.co.uk/2016/11/16/mundo-for-messenger/>.
- Bolter, J., and R. Grusin (2000). *Remediation: Understanding New Media*. Cambridge MA & London, England: MIT Press.
- Brynjolfsson, E., and A. McAfee (2014). *The Second Machine Age. Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York, London: WW Norton & Company.
- Castells, M. (2009). *Communication Power*, Oxford University Press: Oxford.
- Cochran, A. (2016). *Interview*, September 15, 2016, Toronto, CA.
- Cushion, S. (2012). *The Democratic Value of News: Why Public Service Media Matter*, Houndmills. UK: Palgrave MacMillan.
- Davenport, T. (2014). *Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities*. Boston MA: Harvard Business Review Press.
- Digital Advertising Alliance (2015). *Digital Advertising Alliance Announces Mobile Privacy Enforcement to Begin September 1, 2015*. Retrieved January 23, 2017 from <http://digitaladvertisingalliance.org/press-release/mobile-privacy-enforcement>.
- Ess, C. (2009). *Digital Media Ethics: Digital Media and Society Series*. Cambridge UK and Malden MA: Polity Press.
- European Commission (2016). *Reform of EU Data Protection Rules*. Retrieved April 7, 2016 from http://ec.europa.eu/justice/data-protection/reform/index_en.htm.
- Freedman, D. (2014). *The Contradictions of Media Power*. London and New York: Bloomsbury Press.
- Gloor, P. (2006). *Swarm Creativity: Competitive Advantage through Collaborative Innovation Networks*. Oxford: Oxford University Press.
- Głowacki, M., and L. Jackson (eds.) (2014). *Public Media Management for the Twenty-First Century: Creativity, Innovation and Interaction*. London and New York: Routledge.
- Głowacki, M., and L. Jackson (2016). *The Significance of Fluidity in the Network: Organisational Structures within Digital Clusters*. Retrieved January 21, 2017 from <http://ripeat.org/library/2016/7053-significance-fluidity-network-age-organisational-structures-within-digital>.

- Graham, M., and S. De Sabbata, (2013) *The Age of Internet Empires*, Retrieved June 2, 2016 from <http://geography.oii.ox.ac.uk/?page=age-of-internet-empires>
- Hurrell, B., and A. Leimdorfer (2015). *The Data Journalism Handbook*. Retrieved June 20, 2015 from http://datajournalismhandbook.org/1.0/en/in_the_newsroom_1.html.
- Internet Advertising Bureau (2009). *IAB Handbook: A Guide to Online Behavioural Advertising*. Retrived January 23, 2017 from https://iabuk.net/sites/default/files/OnlineBehaviouralAdvertisingHandbook_5455.pdf.
- Iyengar, S., and D.R. Kinder (2010). *News That Matters: Television and American Opinion* (Updated Edition). Chicago: The University of Chicago Press.
- Jenkins, H., S. Ford, and J. Green (2013). *Spreadable Media: Creating Value and Meaning in a Networked Culture*, New York and London: New York University Press.
- Lievrouw, L., and S. Livingstone (2002). *The Handbook of New Media*. London: Sage Publications.
- McKinsey (2015). *An Executive's Guide to Machine Learning*. Retrieved June 3, 2016 from <http://www.mckinsey.com/industries/high-tech/our-insights/an-executives-guide-to-machine-learning>.
- MacCormick, J. (2012). *9 Algorithms That Changed the Future: The Ingenious Ideas That Drive Today's Computers*. Princeton NJ and Woodstock Oxfordshire: Princeton University Press.
- Manovich, L. (2013). *Software Takes Command*. London: Bloomsbury Academics.
- Meier, P. (2015). *Digital Humanitarians: How BIG DATA is Changing the Face of Humanitarian Response*, London and New York: CRC Press.
- Merriam-Webster.com (2016). *Algorithm*. Retrieved January 1, 2017 from <http://www.merriam-webster.com/dictionary/algorithm/>.
- Miller, P. (2010). *Smart Swarm*. London: Collins.
- Pariser, E. (2012). *The Filter Bubble: What the Internet is Hiding From You*. London: Penguin Books.
- Provost, F., and T. Fawcett (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. Sebastopol CA: O'Reilly.
- Pyle, D., and C. San Jose (2015). *An Executive's Guide to Machine Learning, McKinsey Quarterly: June, 2015*. Retrieved June 29, 2015 from http://www.mckinsey.com/Insights/High_Tech_Telecoms_Internet/An_executives_guide_to_machine_learning?cid=Digital-eml-alt-mkq-mck-oth-1506.
- Rogers, S. (2015). Why Data Journalism is the new Punk. In: T. Felle, J. Mair, and D. Radcliffe (eds.), *Data Journalism: Inside the global future*. Bury St Edmonds, UK: Abramis Academic Publishing.
- Siegel, E. (2013). *Predictive Analytics: The Power to Predict who will Click, Buy, Lie, or Die*. Hoboken, New Jersey: John Wiley and Sons.
- Silverstone, R. (2007). *Media and Morality: On the Rise of the Mediapolis*. London: Polity Press.
- Skeggs, B., and S. Yuill (2016). *Values and Value*. Retrieved 15 January, 2017 from <https://values.doc.gold.ac.uk/about/>.
- Solon, O., and J. Wong (2016). *Facebook's Plan to Tackle Fake News Raises Questions over Limitations*. Retrieved December 19, 2016 from <https://www.theguardian.com/technology/2016/dec/16/facebook-fake-news-system-problems-fact-checking>.

- Stuart, D. (2014). *The Marketers Guide to Beacons and Other Location-based Services*. Retrieved April 13, 2016 from <http://digitalmarketingmagazine.co.uk/articles/the-marketers-guide-to-beacons-and-other-location-based-services/937>.
- Tanenbaum, A., and D. Wetherall (2014). *Computer Networks* (fifth edition). Harlow, UK: Pearson Education Limited.
- Vitale, C. (2014). *Networkologies: A Philosophy of Networks for a Hyperconnected Age – A Manifesto*. Winchester UK & Washington USA: Zero Books.
- Williams, O. (2015). Google adviser calls on tech giant to create a separate organisation for data, *The Guardian*, 25 March, 2015. Retrieved July 6, 2015 from <http://www.theguardian.com/media-network/2015/mar/25/google-sandy-pentland-separate-organisation-data>.